information

MDPI

Article

FPGA-Based Voice Encryption Equipment under the Analog Voice Communication Channel

Xinyu Ge, Guiling Sun *, Bowen Zheng and Ruili Nan

College of Electronic Information and Optical Engineering, Nankai University, Tianjin 300350, China; 2120190289@mail.nankai.edu.cn (X.G.); zhengbwen@mail.nankai.edu.cn (B.Z.); 1120200133@mail.nankai.edu.cn (R.N.)

* Correspondence: sungl@nankai.edu.cn

Abstract: This paper describes a voice encryption device that can be widely used in civil voice call encryption. This article uses a composite encryption method to divide the speech into frames, rearrange them in the time domain, and encrypt the content of the frames. The experimental results show that the device can complete the encryption normally under various analog voice call conditions, and the voice delay, quality, encryption effect, etc. are guaranteed. Compared with traditional time-domain encryption, it effectively solves the original voice information remaining in the encrypted information, and further increases the security of the voice.

Keywords: FPGA; voice source real-time encryption; speech scrambling

Citation: Ge, X.; Sun, G.; Zheng, B.; Nan, R. FPGA-Based Voice Encryption Equipment under the Analog Voice Communication Channel. *Information* **2021**, *12*, 456. https://doi.org/10.3390/info12110456

Academic Editor: Willy Susilo

Received: 6 October 2021
Accepted: 3 November 2021
Published: 4 November 2021

1. Introduction

The purpose of voice communication is to realize the transmission of information in human daily life, and people's requirements for voice calls are increasing continuously and developing rapidly [1]. Voice communication changes people's communication methods; because of its immediacy, convenience and other characteristics, it is widely used, and it is the main method in modern communication systems [2]. With the development of communication technology, the security of traditional voice communication has been greatly challenged, and a series of voice signal eavesdropping cases have occurred [3]. Party and government organizations in various countries attach great importance to the confidentiality of voice communications. Enterprises hope that voice communications will not compromise their confidentiality, and individuals hope to protect their privacy [4].

Encrypted voice communication is attracting people's attention. The militaries of various countries have established dedicated lines for encrypted communication, but the cost is high and the applicability is poor. In recent years, many companies have introduced voice encryption equipment, but they often require additional communication equipment, which is costly and cannot be integrated into existing communication equipment or lines. Common encryption methods are generally divided into digital encryption and analog encryption. Generally speaking, digital encryption is more reliable, but due to bandwidth and other restrictions, voice encryption generally uses analog encryption [5]. Analog voice encryption is also called "speech scrambling". Speech scrambling techniques are some of the methods used to prevent informatics attacks and to protect the information from intruders or interceptors. This can ensure the security of voice transmission [6].

This article introduces an encryption device for real-time voice communication based on the FPGA design [7,8]. It does not require special lines or additional voice call equipment. It combines voice scrambling and FPGA digital signal processing to produce a hybrid encryption method that is suitable for various types of voice communication [9]. The appearance of this product is just like a headset, and it can be used after plugging in a mobile phone. The voice signal is collected and encrypted by the encryption device, and then sent to voice communication devices such as mobile phones for transmission through

the communication channel, which solves the security problem of eavesdropping during the voice transmission process and protects the privacy of users.

The rest of this article is organized as follows: Section 2 introduces the application of voice encryption equipment to analog voice communication channels. Section 3 describes the system architecture. Section 4 tests the system and analyzes the results. Section 5 summarizes the article.

2. Materials and Methods

Voice encryption is generally divided into digital encryption and analog encryption. Digital encryption first samples the voice signal, then encrypts and modulates it, modulates it on the carrier and sends it out. Using digital modulation inevitably increases the bandwidth of the call, so exceeding the frequency band of the general voice call means, thus it is rarely used in end-to-end voice communication encryption [10].

Voice-like modulation is an encryption method suitable for the transmission of information in the voice channel, which retains the voice characteristics of the signal and reduces the interference of the vocoder on the signal. However, most of its current modulation technologies are based on the realization of limited ideal environments. If multiple vocoder conversions, actual channel errors, packet loss and other uncertain factors are considered, there is no mature technology to choose from [11,12].

Analog encryption is generally called voice scrambling. In this method, there is no need to digitize and compress voice samples, nor to use modulation, demodulation, and demodulation. For analog encryption, common encryption methods include: time domain encryption, frequency domain encryption, amplitude encryption, and composite encryption.

In order to achieve a secure encryption, and a real-time voice encryption system that is not restricted by type of mobile phone, walkie-talkie, or communication network, the technical means that can be adopted are generally frequency domain encryption. At present, the commonly used frequency domain encryption generally works by moving the audio spectrum of the call, i.e., moving the high frequency part to the lower part of the frequency band and moving the low frequency part to the upper part of the frequency band. The original voice signal is converted into an incomprehensible sound, so as to achieve the purpose of encryption. Frequency domain encryption can retain the characteristics of the voice signal, and is easy to implement and reproduce. But its disadvantage is that the encryption is relatively simple, retains the voice call characteristics, and is easy to crack.

Time domain encryption usually divides the speech signal into frames; each frame contains several samples. Afterwards, these frames are rearranged, and the speech signal is scrambled through the key to the arrangement process. Speech scrambling changes intelligible speech into incomprehensible speech, so that the eavesdropper cannot understand the original meaning. Since time domain encryption itself is processed in the analog domain, the signal amplitude remains continuous and the frequency does not change much (less than or equal to 4 kHZ), and the reserved voice features are similar to ordinary voice signals, which can be more easily passed through the vocoder and communication channel. Time domain encryption retains much original information about the voice, and has significant shortcomings in terms of its security and intelligibility [13].

In order to solve the above-mentioned shortcomings of time-domain encryption, the system we have designed divides the voice data into frames and stores them in the SDRAM. Then, the positions of each frame are rearranged according to time-domain encryption, and the synchronization signal is used to determine the encryption period, which further increases the security of the encrypted signal and reduces its intelligibility.

3. System Structure

This article describes an FPGA-based voice encryption system, and the system architecture is shown in Figure 1. We only need this device to encrypt the signal between headphone and mobile device. The following describes the product system structure in detail from the chip selection and PCB design.

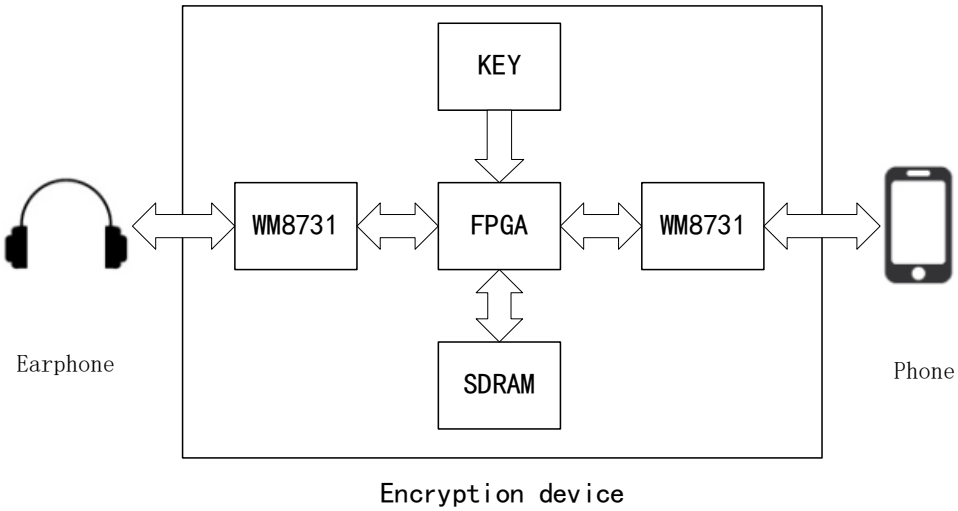


Figure 1. Voice encryption system architecture.

3.1. Hardware Selection and Design

3.1.1. Audio Chip

This system uses the WM8731 chip, which is used for the A/D and D/A conversion of the sound signal in the process of acquisition and playback. The input clock is generated from within the FPGA and sent to the chip. The initialization of the chip and the working status and functions of the chip are realized by configuring the 11 internal registers corresponding to the I2C bus. The sampling frequency (A/D and D/A sampling rate) of the WM8731 chip in this system is set to 48 kHZ, and the converted data bit length is 16 bits [14].

Two WM8731 chips are used in this system, which respectively encrypt and decrypt the sound information from the headset and mobile phone. There are two WM8731 chips, and the internal registers are set separately through the I2C bus to achieve different functions. The WM8731 chip includes two input modes, the line input and microphone input, all of which are designed for the line input in this system. The WM8731 chip on the earphone side needs to turn on the microphone gain to amplify the tiny signal of the microphone to proceed to the next step. The WM8731 chip on the mobile phone side does not need to be amplified, and the microphone gain needs to be turned off. In addition, when inputting and outputting sound signals, different gain attenuations need to be adjusted accordingly.

The digital audio interfaces of the WM8731 chip are: the digital audio clock BCLK, the digital audio input DACDAT, the DAC left and right channel acquisition clock DACLRC, the digital audio output ADCDAT, and the ADC left and right channel acquisition clock ADCLRC.

The incoming voice information becomes digital information after being collected by the WM8731 chip. The DACDAT, DACLRC, ADCDAT, and ADCLRC collect the data and transmit on each falling edge of the BCLK. As shown in Figure 2, the BCLK is the sampling clock, and each clock cycle transmits one bit of data. The DACLRC and ADCLRC control the collection of the left and right channels. The left channel is collected during the rising edge and the right channel is collected during the falling edge. The acquisition cycle is 64 BCLK clocks, and 64-bit data, each of which occupies 32 bits for the left and right channels [15].

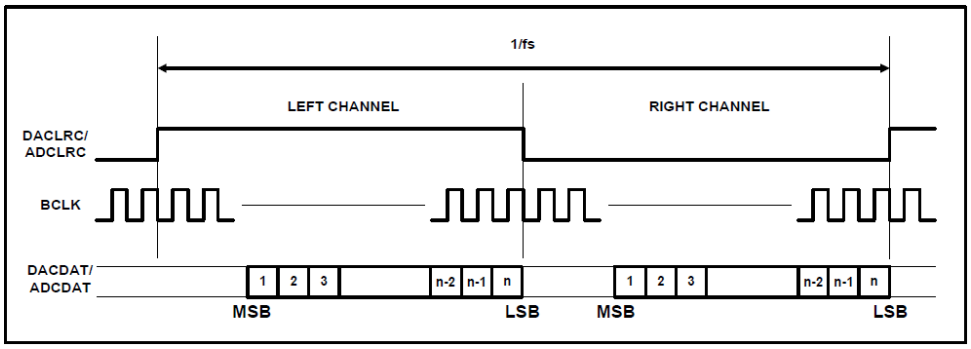


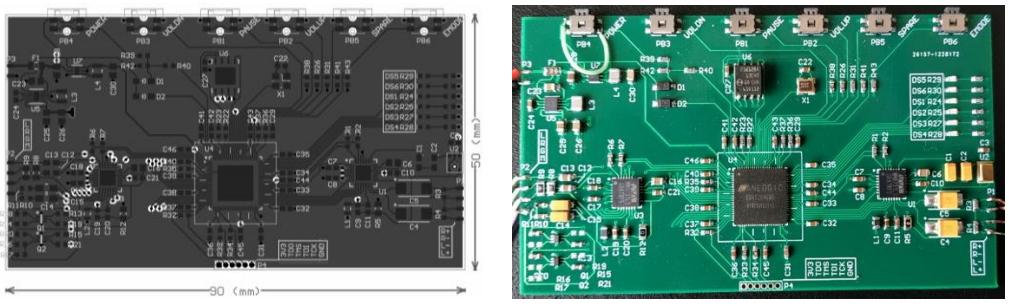
Figure 2. WM8731 chip in the right justified mode.

3.1.2. FPGA Chip

The FPGA chip uses Anlogic Technologies' EG4S20 as the core unit. It has 20 K logic cells (LUT4/LUT5 hybrid architecture), about 130 KB of SRAM (Static Random Access Memory), a built-in 32-bit wide 64-MBit SDRAM (Synchronous Dynamic Random Access Memory), rich LVDS (Low-Voltage Differential Signaling) pins, and a built-in 12-bit 1MSPS (Million Samples per Second) ADC (Analog to Digital Converter). Compared with the Altera and Xilinx chips, it is cheap and integrates the SDRAM. It takes up less space, which allows us to make the entire device smaller. In summary, we used the EG4S20 chip from Anlogic Technologies [16].

3.1.3. PCB Design

The design of the circuit board of this system is shown in Figure 3. The left side of the product is the earphone soldering point, and the right side is the WM8731 chip, which converts the data from the FPGA into an audio signal and sends it to the earphone. There is a patch microphone on the top to provide input sound. On the upper side are 6 LED indicators, which are used to indicate the working status of the product. Located in the middle is the FPGA chip, which mainly performs data encryption and decryption algorithms. On the left are the soldering points of the four-ring earphone cable and the WM8731 chip, which are used to receive the data from the mobile phone and transfer it to the FPGA. At the top of the product are six key control buttons to switch the product on and off, adjust the volume, and so on. The bottom of the product is used for debugging and downloading data.



(**a**) (**b**)

Figure 3. The voice encryption equipment's PCB design (**a**) and physical diagram (**b**).

3.2. Synchronization Algorithm

When encrypting the voice signal, a certain degree of clock synchronization is required. Because the encryption needs to divide the signal into frames in the time domain, it is necessary to synchronize the clock to determine the starting position of the frame when decrypting. Compared with the clock of digital communication in general, the clock of this product does not need very strict alignment. When performing voice collection, several

adjacent points will all have the same magnitude of sound amplitude, and the collected data are almost the same. Therefore, even if the clock is identified with a certain error, it will not affect the decryption of the product. During the actual testing of this product, the clock difference generally did not exceed 10 sampling points, which is about 10/≈0.2 ms, and so does not affect the encryption and decryption at all. Thus, in the product design process we can simply design the synchronous clock of the product [17].

This article adopts a synchronization method to calculate the number of fixed sine waves: first, a sine wave with a fixed frequency and a fixed number of cycles is generated by the transmitter. The frequency of the sine wave needs to be within the range of the human voice, preferably 1000–2000 Hz. It is sent separately by the encryption equipment of both parties. The encryption and decryption in this product is not performed synchronously, but the sender's encryption clock is the same as the receiver's decryption clock, so both parties need to send a piece of synchronization information. In one device, encryption and decryption are performed in its two clocks. The encryption and decryption is performed by the WM8731 audio chip.

After the receiver receives the synchronization information, it starts to generate the corresponding synchronization clock. First of all, the synchronization clock cycle is agreed upon, and is a fixed frequency. It only needs to know when the clock starts. As shown in Figure 4, the receiving end reads the number of sine waves from the synchronization signal. When the count of sine waves reaches the predetermined number to be sent, the clock is generated, and the RST bit of the clock is activated to make the signal regenerate from its beginning at this time, so that we get the same clock as the original synchronization signal.

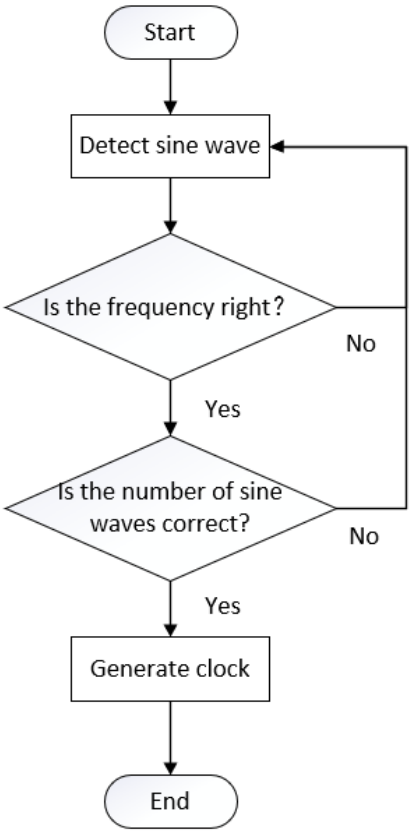


Figure 4. Synchronous clock generation.

In the process of transmission, the synchronization signal passes through the voice channel and is inevitably affected by noise. Calculating the number of sine wave cycles to determine the clock can greatly reduce the impact of noise. We can count the number of times at each peak of the sine wave, but due to the influence of noise, we are likely to

have the following situation; namely, the noise interferes with the detection of the peak, so that it is impossible to count accurately. Therefore, in order to recover the clock more accurately, we also added a hysteresis comparator to filter the signal in the process of signal processing. We set a threshold amplitude. When the transition amplitude is not greater than the threshold, the output voltage value is stable. When the transition amplitude exceeds the threshold, the reverse voltage is output to ensure that the synchronization signal is not caused by noise. The number of sine waves is incorrectly calculated due to the influence of noise, resulting in synchronization failure.

The synchronization signal needs to be sent before the voice call. After the call is connected, the sender needs to manually press the synchronization button to send a synchronization signal, so that a synchronization is completed, and the call can be conducted normally. As long as the encryption devices on both sides are not powered off, the clocks will always be synchronized.

3.3. Encryption Algorithm

This device is designed with an in-built encryption algorithm, and the voice signal is transformed into ciphertext under the encryption algorithm for transmission in the channel to protect the privacy of the user. The following details how to implement the in-built encryption algorithm.

The voice signal passes in and out of the headset, microphone, and mobile phone to the WM8731 chip. The WM8731 chip samples the data once every lrc clock, quantize it, and converts it into 32-bit binary data left_data and right_data. For this product, two-channel microphones are not used, so the data of the left and right channels are the same, and thus only one side of the data needs to be processed.

For each datum collected by the lrc clock, an address generation module assigns addresses to it in turn, and stores them in the SDRAM that comes with the chip. Because these addresses are stored in accordance with the specified address generation rules, they are used later. It is much more convenient when the FPGA chip encrypts the data.

The voice data collected by the WM8731 audio chip are stored in the SDRAM according to the address number, grouped according to certain rules, and then encrypted by the FPGA. The voice signal becomes 64-bit data after sampling by the WM8731 audio chip, and the sampling frequency is 48 kHZ. The collected voice data is stored in the SDRAM according to the address number, and every n adjacent datum forms a frame, where n is the frame size. The m frames are combined into an encryption period, where m is the number of frames in a period. Under normal circumstances, people's speaking speed per minute is roughly 100–200 words per minute, with an average of about 2–4 words per second, or about 5–10 syllables. When we shorten the time to 50 ms or even lower [18], it is difficult for us to understand what the sound is. The WM8731 chip in this system works at a sampling rate of 48,000, which means that samples are taken 48,000 times per second. We took 0.4 s as an encryption cycle, which contains 19,200 sampling points, and then divided the 19,200 sampling points into 40 encrypted information segments on average. Each encrypted information segment contains 480 sampling points and takes 10 ms. For every 40 ms of sound, people cannot hear what the starting sound is. As shown in Figure 5, our device shuffles the order of these 40 encrypted information segments, destroys the voice information contained in the original sound, and protects its important content from eavesdropping. The receiver can recover the original voice information according to the pre-arranged scrambled order [19,20].

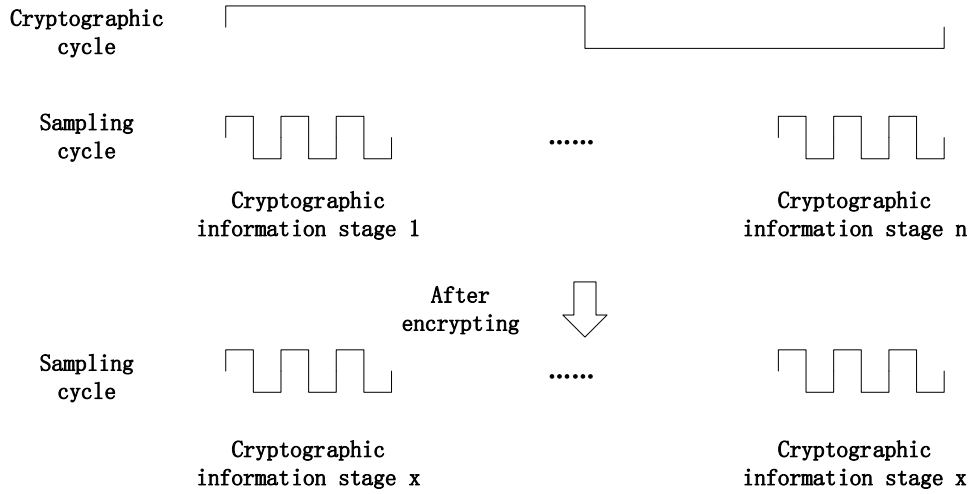


Figure 5. Disrupting the order of encrypted information segments.

For the rearrangement method, we adopted the following methods:
Input: n: The number of frames in a period
X: Sound data for each frame
Output: Y: Scrambling sound data for each frame
Step 1: set k = 1
Step 2: loop until k = n/2
Step 2.1: set Y(k) = X(k)
set Y(2k) = X(k + 1)
Step 3: return Y

In addition, before rearranging the time frames, the data in each frame is also encrypted; that is, the data stored in the SDRAM first is taken out last, and the subsequent data is taken out first. The whole process is completed by the address control module. The data address written into the SDRAM runs from low to high, and the read address runs from high to low. In this way, the reverse reading of the data is completed, and the voice security is further ensured.

4. Test Results and Analysis

In order to verify the effectiveness of the product design in this study for encrypted calls, we conducted the following tests on the voice encryption effect when using mobile phones for calls:

4.1. Time Delay Analysis

The collected voice data needs to be encrypted inside the device, and continues to be sent after the encryption is completed, so the encrypted device causes a certain delay. The most critical factor affecting the delay is the size and number of frames, because the encrypted information only continues to be sent after all the data processing in an encryption cycle is completed. For real-time voice call devices, the delay of voice greatly affects people's voice experience. When the delay was within 150 ms at that time, the sound heard by the human ear was found to be very smooth and there was no difference; when the delay reached 150 ms–400 ms, the voice heard by people was delayed, but it did not affect the normal call [21]; the delay was too high at that time. When the delay exceeds 1000 ms, the other party's voice is slow during the call, which affects the normal call effect. The delay of our product is consistent with the encryption cycle, and the delay can be changed by adjusting the encryption cycle [22]. In a call environment that requires a good experience, the encryption cycle can be shortened to obtain a good call experience. The shortest possible encryption period is 0.1 s. If it continues to be shortened, the change in the audio frequency range increases, and the distortion is more obvious. If the total

encryption time is too short, then there is not much difference in the sound of the same tone, resulting in an unsatisfactory encryption effect. If one does not pay much attention to the voice delay and requires more secure confidentiality, one can appropriately lengthen the encryption period, extend the encryption time to 0.4 s, and increase the number of encrypted information segments, which can greatly increase the difficulty of deciphering and increase the confidentiality of the product [23]. The frame size and the number of frames depend on the required encryption strength and delay. Increasing the frame size and the number of frames can effectively increase the encryption strength, but it will increase the delay of the product. The reasonable allocation of frame sizes and frame periods is something that must be considered when setting the encryption parameters. Table 1 below shows the received voice delay time under different encryption parameters:

Table 1. Voice delay time test.

Group	Frame Length	Number of Frames	Time Delay/ms
1	600	4	50
2	600	8	100
3	1200	8	200
4	1200	16	400
5	2400	4	200
6	2400	8	400

4.2. Voice Quality Test

Because this design is based on the time sequence segmentation method, the voice information is encrypted, and the waveform of the voice signal is retained to the greatest extent, so that the loss of voice during the channel transmission and through the vocoder is reduced to a minimum. However, due to the cutting of the voice signal, a sudden change occurs in the voice signal that is continuous, and great changes occur in the frequency domain, which exceed the frequency upper limit of the voice channel transmission, causing the signal to be distorted. At that time, voice information that is different from the original signal is inevitable [24]. In order to solve the above problems, we cannot cut the signal too short, so as not to cause a large number of mutations in the signal. People are not sensitive to small voice errors, and these do not affect normal voice calls. At the same time, because people's speech pitches are generally within a relatively stable range with little fluctuation, the difference in sound loudness minimizes the impact on signal cutting and encryption. However, for the sounds of music that has great changes in tones and requires high sound quality, the sound quality of speech is more affected [25].

For the subjective analysis of the voice call quality, this study invited some people to try the device and fill out a questionnaire to objectively evaluate the voice call quality. As shown in Table 2, five people were invited to this subjective test to score the voice call quality, with a full score of five. The results are shown in the Table 2. Everyone could hear the original voice message clearly and the encryption did not affect the normal call; a small number of people thought that there was still a certain amount of noise and the call quality was not good enough, but it did not affect the call.

Table 2. Voice call quality test.

Tester ID	Voice Quality	Noise
A	5	None
B	4	Little
C	4	Little
D	3	Noisy
E	4	Little

In addition, the design has a wide range of applications and strong portability. For various voice call scenarios, such as voice call channels in communication apps and walkie-talkies, it offers good call quality.

4.3. Encryption Strength Analysis

First of all, for the effect of anti-eavesdropping, during the transmission of the voice signal, the intercepted signal is an encrypted signal, and the original voice cannot be heard directly due to the segmentation and scrambling [26]. For greater encryption strength, the encryption cycle is divided into several encrypted information segments, and then scrambled, which greatly increases the encryption strength. As shown in Table 3, since the encryption period can also be dynamically adjusted from 0.1 s to 1 s, the number of encrypted information segments can also be adjusted according to need.

Table 3. Voice call encryption parameters range and step size.

Encryption Parameters	Range	Step Size
Encryption period	0.1–1 s	None
Number of frames	4–40	1
Frame size	200–2000	10
Frame rearrangement	Determined by different arrangements	

If we want to decipher encrypted speech, we need to know the length and number of encrypted frames. When the cracker does not know the detailed parameters, he needs to constantly test the length and number of frames one by one. The size of frame length selected was from 200 to 2000 and the number of frames selected was from 4 to 40. This contains up to 6480 parameter choices. Furthermore, there are various frame arrangements, each parameter can have 20 arrangements, and there can be about 12,960 encryption modes in total.

In terms of hardware, when the cracker does not have a voice encryption device, he also needs to build an FPGA development board with two WM8731 audio processing chips and debug the device to meet the call and decryption needs.

In terms of software, to crack the encryption, the above parameters need to be compiled into a suitable program. Each time one tests the parameters, one needs to recompile and download. In actual testing, it took at least 1 min to complete each time one downloaded the program and tested it, which effectively increased the need for cracking. Without calculating the time required for software and hardware development, the time required to test all parameters should be at least 200 h.

When the encryption parameters are leaked, we can also easily change the parameter selection of the device. By regularly changing the internal parameters of the encryption device, we can effectively prevent the leakage of important voice information.

Compared with other analog voice encryption devices, this device divides the voice into frames and rearranges them, and encrypts the content of the frames, effectively solving the large amount of voice information remaining in traditional voice encryption. By changing the parameters, the encryption strength under the analog voice channel is greatly improved. Although traditional digital encryption has a higher encryption strength than this device, it is not suitable for analog voice channel calls due to factors such as bandwidth.

5. Conclusions

Aiming at the security problem of real-time voice encryption in mobile communication systems, this paper proposes a voice encryption system based on dynamic timing cutting and scrambling, which encrypts the signal before the voice channel is transmitted. This article uses the Anlogic chip as the core, and the WM8731 audio chip for sound collection and generation, which effectively reduces the cost, facilitates mass production, and enables the device to be more widely used in commercial and civilian applications. The system provides effective solutions to the noise of the voice channel, the limited bandwidth, and

the interference of the vocoder, ensuring that the receiving end can receive the correct, low-latency, and highly confidential voice signal. The test results show that the design can be used normally, whether it is the mobile phone call channel, the voice channel of various software apps, or the call channel of the walkie-talkie, and it has a wide range of applicability. The design of the product also has some flaws: in order to balance the delay and encryption effects, either the call experience or the degree of security is affected; because the cutting in the time domain destroys the sound, the restored sound contains some defects. Although the bottom noise does not affect the call, it still brings a bad user experience.

Author Contributions: Data curation, R.N.; funding acquisition, G.S.; software, X.G.; validation, B.Z.; writing—review & editing, X.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Tianjin Science and Technology Major Project and Engineering, grant number No.18ZXRHNC00140

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Park, S.; Cho, H.; Park, Y.; Choi, B.; Kim, D.; Yim, K. Security problems of 5G voice communication. In Proceedings of the 21st International Conference on Information Security Applications, WISA 2020, Jeju Island, Korea, 26–28 August 2020; pp. 403–415.
2. Sadkhan, S.B.; Al-Sherbaz, A.; Mohammed, R.S. Chaos based Cryptography for Voice Encryption in Wireless Communication Literature Survey. In Proceedings of the 1st International Scientific Conference on Electrical, Communication, Computer, Power, and Control Engineering (ICECCPCE), Mosul, Iraq, 17–18 December 2013; pp. 191–197.
3. Helenport, A.; Tait, B.L. Aspects of Voice Communications Fraud. In Proceedings of the 11th International Conference on Global Security, Safety, and Sustainability (ICGS3), London, UK, 18–20 January 2017; pp. 69–81.
4. Chen, Y.; Hao, J.; Chen, J.; Zhang, Z. End-to-end speech encryption algorithm based on speech scrambling in frequency domain. In Proceedings of the 3rd International Conference on Cyberspace Technology, CCT 2015, Beijing, China, 17–18 October 2015.
5. Sadkhan, S.B.; Salah, A. The trade-off between security and quality using permutation and substitution techniques in speech scrambling system. In Proceedings of the 2019 First International Conference of Computer and Applied Sciences (CAS), Baghdad, Iraq, 18–19 December 2019; pp. 244–249.
6. Enache, F.; Deparateanu, D.; Oroian, T.; Popescu, F.; Vizitiu, I. Theoretical and practical implementation of scrambling algorithms for speech signals. In Proceedings of the 7th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2015, Bucharest, Romania, 25–27 June 2015; pp. S49–S52.
7. Huang, H.H.; Liu, J.R.; Wang, Y.D. The Design Of Bluetooth Voice Source Wireless Real-time Voice Encryption System. In Proceedings of the 2016 3rd International Conference on Materials Engineering, Manufacturing Technology and Control, Taiyuan, China, 27–28 February 2016.
8. Riyadi, M.A.; Pandapotan, N.; Khafid, M.R.A.; Prakoso, T. FPGA-based 128-bit Chaotic Encryption Method for Voice Communication. In Proceedings of the International Symposium on Electronics and Smart Devices (ISESD)-Smart Devices for Big Data Analytic and Machine Learning, Bandung, Indonesia, 23–24 October 2018; pp. 34–38.
9. Zhang, Y.P.; Duan, F.; Liu, X. The Research of Applying Chaos Theory to Speech Communicating Encryption System. In Proceedings of the International Conference on Multimedia, Software Engineering and Computing, Wuhan, China, 26–27 November 2011; pp. 197–202.
10. Abro, F.I.; Rauf, F.; Chowdhry, B.S.; Rajarajan, M. Towards Security of GSM Voice Communication. *Wirel. Pers. Commun.* **2019**, *108*, 1933–1955. [CrossRef]
11. Islam, S.; Haq, I.U.; Saeed, A. Secure end-to-end SMS communication over GSM networks. In Proceedings of the 2015 12th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Islamabad, Pakistan, 13–17 January 2015; pp. 286–292.
12. Rehman, M.U.; Adnan, M.; Batool, M.; Khan, L.A.; Masood, A. Effective Model for Real Time End to End Secure Communication Over GSM Voice Channel. *Wirel. Pers. Commun.* **2021**, *119*, 1643–1659. [CrossRef]
13. Pal, P.; Sahana, B.C.; Ghosh, S.; Poray, J.; Mallick, A.K. Voice Password-Based Secured Communication Using RSA and ElGamal Algorithm. In Proceedings of the 5th International Conference on Advanced Computing and Intelligent Engineering, ICACIE 2020, Université des Mascareignes (UdM), Beau Bassin-Rose Hill, Mauritius, 25–27 June 2020; pp. 387–399.

14. Zhong, W.; Wang, Y. Interfaces design of dual-channel audio codec based on FPGA. *Int. J. Adv. Comput. Technol.* **2011**, *3*, 460–466. [CrossRef]
15. Yuan, H.; Xu, P. Designing of the digital voice recording system on SOPC. In Proceedings of the 2012 IEEE Fifth International Conference on Advanced Computational Intelligence (ICACI), Nanjing, China, 18–20 October 2012; pp. 1213–1215.
16. Holmberg, W. Cost-Efficient Method Forlifetime Extension Ofinterconnectedcomputer-Based Systems. Available online: https://www.diva-portal.org/smash/get/diva2:1600354/FULLTEXT01.pdf (accessed on 2 November 2021).
17. Ma, H.; Wang, Y.; Li, G. Implementation of audio data packet encryption synchronization circuit. In Proceedings of the 1st Euro-China Conference on Intelligent Data Analysis and Applications, ECC 2014, Shenzhen, China, 13–15 June 2014; pp. 321–329.
18. Shaofeng, L.; Chaoping, G.; Lin, N.; Wanli, K.; Minjiao, Z. The Research of Encryption Algorithm for Voice Communication of Mobile Station. In Proceedings of the 2015 International Conference on Intelligent Transportation, Big Data and Smart City, Halong Bay, Vietnam, 19–20 December 2015; pp. 898–901.
19. ElChabb, R.; Khattar, F.; Bassoul, G.; ElMurr, S.; Atallah, J.G. RT-VED: Real Time Voice Encryption/Decryption. In Proceedings of the International Colloquium on Computing, Communication, Control, and Management (CCCM 2010), Yangzhou, China, 20–22 August 2010; pp. 270–273.
20. Sharma, P.; Sharma, R.K. Design and Implementation of Encryption Algorithm for Real Time Speech Signals. In Proceedings of the Conference on Advances in Signal Processing (CASP), Pune, India, 9–11 June 2016; pp. 237–241.
21. Lwin, H.M.M.M.; Ishibashi, Y.; Mya, K.T. Influence of Voice Delay on Human Perception of Group Synchronization Error for Remote Learning: One-way Communication Case. In Proceedings of the 2020 IEEE Conference on Computer Applications, ICCA 2020, Yangon, Myanmar, 27–28 February 2020.
22. Liu, X.; Chen, J.; He, X.; Li, W.; Zhang, G. A study of radio voice signal based on the time delay estimation. *Int. J. Hybrid Inf. Technol.* **2016**, *9*, 103–110. [CrossRef]
23. Peng, Z.; Mo, K.; Zhu, X.; Chen, J.; Chen, Z.; Xu, Q.; Ma, X. Understanding user perceptions of robot's delay, voice quality-speed trade-off and GUI during conversation. In Proceedings of the 2020 ACM CHI Conference on Human Factors in Computing Systems, CHI EA 2020, Honolulu, HI, USA, 25–30 April 2020.
24. Lubkowski, P.; Polak, R.; Sierzputowski, R. The measurements of the secured voice communication quality in a broadband radio channel. In Proceedings of the Radioelectronic Systems Conference (RSC), Jachranka, Poland, 20–21 November 2019.
25. Yihunie, F.; Abdelfattah, E. Simulation and Analysis of Quality of Service (QoS) of Voice over IP (VoIP) through Local Area Networks. In Proceedings of the 9th IEEE Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2018, New York, NY, USA, 8–10 November 2018; pp. 598–602.
26. Cox, R.V.; Bock, D.E.; Bauer, K.B.; Johnston, J.D.; Synder, J.H. Analog Voice Privacy System. *ATT Tech. J.* **1987**, *66*, 119–131. [CrossRef]